

Stochastic Algorithms in the Large: Batch Size Saturation, Stepsize Criticality, and Exact Dynamics

Courtney Paquette

joint work with Kiwon Lee, Andrew Cheng, Elliot Paquette

Workshop in honor of the 60th Birthday of Adrian Lewis
Seattle, 9–10 August, 2022

In this talk, I will present a framework, inspired by random matrix theory, for analyzing the dynamics of stochastic optimization algorithms (e.g., stochastic gradient descent (SGD) and momentum (SGD + M)) when both the number of samples and dimensions are large. Using this new framework, we show that the dynamics of optimization algorithms on a least squares problem with random data become deterministic in the large sample and dimensional limit. In particular, the limiting dynamics for stochastic algorithms are governed by a Volterra equation. From this model, we identify a stability measurement, the implicit conditioning ratio (ICR), which regulates the ability of SGD+M to accelerate the algorithm. When the batch size exceeds this ICR, SGD+M converges linearly at a rate of $O(1/\sqrt{\kappa})$, matching optimal full-batch momentum (in particular performing as well as a full-batch but with a fraction of the size). For batch sizes smaller than the ICR, in contrast, SGD+M has rates that scale like a multiple of the single batch SGD rate. We give explicit choices for the learning rate and momentum parameter in terms of the Hessian spectra that achieve this performance. Finally we show this model matches performances on real data sets.